达尔文是干什么的，以及语言、教育和智力研究

作者：顾立帆

研究智力应该从生命开始。生命和智力科学的研究也跟传统的学术和成功术有很大不同。达尔文并非传统意义上的栋梁之才，其成功需要特别的环境。

亚里斯多德是生物分类学家，孟德尔是遗传学家。达尔文是干什么的？准确一点说是演化学家。演化学主要是生物历史。也就是说达尔文主要是生物历史学家。

但演化学不仅仅是生物历史研究，也是生物科学研究。新冠病毒流行后，演化树和证据链成了病毒研究的关注焦点。以后太空宇航大规模发展起来后，在高辐射、重力多变等等更复杂的环境里，演化会更加快速和复杂。也就是说，达尔文的演化理论也会越来越重要。

所以，未来太空探索和发展远比在新大陆建立殖民地复杂及困难，其战略需要根据生命演化和智力发展的特点研究。需要特别避免可能导致严重文明退化的生命演化和智力现象，包括奴役、遗传性畸变以及其它更严重问题，需要更周密的长期研究和发展计划。

而哪些变异会有利于人类在太空的发展，哪些会造成损害，是更加复杂的问题。类似 leukotomy 手术后对智力的那种测试以及那种测试所基于的智力分类，明显不能正确反映对智力的影响，包括对自由意志的影响。需要对智力的系统结构作更准确的研究。

亚里斯多德生物分类法以外在形态为基础，有很多误导。现代生物分类法以达尔文的演化理论为基础，按基因演化历史分类，但有新问题。选择哪些演化分支点作为分类基础非常主观，出现很多不同的分类方法。唯一有严格规范的物种定义要求不同物种间有生殖隔离。而在基因技术突飞猛进的时代，生殖隔离要求恰恰会较容易打破，导致更大的混乱。

生物分类法本来希望成为一种标准，但实际上是语言的选择，结果出现语言的多样化。现在生物分类五花八门，分类方法越来越多。这跟化学元素周期表和量子力学标准模型的定义严谨、层次分明形成明显对照，显示了生命科学跟物理科学有本质区别。物理科学包括经典物理、化学、量子物理等等，生命科学包括生物学、生理医学、心理学等等。

如果生物分类法不得不允许多样化，并非以基因演化为基础的郑作新鸟类分类法是不是还有存在价值？郑作新分类法跟以形态为基础的传统分类法也有些不同，代表一种较典型的语言思维习惯。游禽、涉禽、攀禽、陆禽是较好的生态分类，鸣禽、猛禽却是跟前面四种生态分类不同的典型分类。这种混杂的分类法，也是深度学习的典型特征和问题。

事实上，类似游禽、涉禽、攀禽、陆禽等等的生态模式不是鸟类独有的。生态模式和基因的关系和互动也远比想象的复杂。

生命科学研究中，Molecular Clock 的有效范围也有争议。如果说在物理科学和生命科学中，概率统计方法只是有争议，在智力研究中概率统计方法会产生严重问题则可通过实验验证。缺乏准确的逻辑临界点分析，谈论智力研究正确率有多高没有意义。在模式不稳定演变下，正确率会急剧变化。

深度学习则还有更本质的问题：行为主义（Behaviorism）的缺陷。行为主义本来是心理学的一种学派，强调训练、强化和控制。种植业和制造业的很多工作，大部分人通过训练确实可以做，这是行为主义的基础。但科学研究和其它很多行业的很多工作，大部分人通过训练仍不适合，只有少数有特长的人才能胜任。而且能胜任某些研究的人不一定适合其他研究，这在科学时代成为行为主义的致命缺陷。

控制是动物生理层次的智力，对人类实施控制会造成其思维的降阶和降维等等问题，在某种程度上损害人类特有智力。所以对人类的控制术实际上是一种 tradeoff。新科学的探索需要更多地发挥自由意志的优势，尽可能避免控制术带来的危害。

基于生命的行为主义失败了，但人工智能又给行为主义带来新的狂热：训练、强化和控制真能造就万能的人工智能？有人工智能专家说“人类正在扮演上帝的角色”。Demis Hassabis 则宣称 Deepmind 的技术是 potentially a meta-solution to any problem。Deepmind 最近的观点是 Reward is Enough。这些在学术上并不可能。

“扮演上帝的角色”不但需要准确理解人类智力、准确解释 causation 关系，还需要准确理解上帝的逻辑。人工智能专家能做到吗？Demis

Hassabis 真能解决 any problem，包括裁决信仰争议，解决波函数坍塌、多世界诠释等等量子力学诠释争议和其它各种科学争议?

如果 Demis Hassabis 的技术真能解决 any problem，用该技术不就能建立更优越的普世计划经济模式? 事实上就有一种从垄断资本主义到社会主义的理论。计划经济恰恰就是计算思维的产物，马尔萨斯和罗马俱乐部的人口理论也是（笔者不反对计划经济，只是认为计划经济不是普世模式）。人工智能专家没有意识到这一点，是因为缺乏良好的通识教育。

由此可见他们的语言智力和逻辑判断能力有问题。这也暴露了计算思维和一阶逻辑训练的局限性。事实上，他们的语言处理技术缺乏重要的抽象思维、语义演变、高阶逻辑、诡辩分析和判断力等等方面的有效测试，当然无法准确理解人类智力，因此也无法描述无人汽车技术的本质缺陷。

智力现象比生命演化现象节奏快得多，问题也更严重。自动驾驶标准中的问题会导致的后果当然比生物分类法严重得多。现有标准没有找到关键的有效观察量，也无法有效测试。在模式不稳定演变下会出现严重问题。根据自由意志的研究就可以证明 SAE level 4 自动驾驶不可能实现，当然还有更严重的问题。

好几年前听说无人汽车技术已经成熟，只是法律落后；后来发现并非如此。所以 Waymo 无人汽车到底有没有通过 SAE level 4 测试，依据是什么? 应该给公众严谨回答，以便进一步研究。如果说在某些条件下通过了，应该仔细描述通过的条件。笔者可以证明这些条件无法准确描述并有效控制。而证明 SAE level 5（即全自动驾驶）是不可能的而且会遭遇大概率问题，则相对容易些。

按媒体报道，目前无人汽车 SAE level 4 的测试实际上是按地区测试并按地区宣布通过测试的，也就是实际上在测试中分成一些典型的类型如：亚利桑那人口密度较低、气候晴朗干燥地区，旧金山人口密度较高、雨季多雾多雨地区，以及更复杂的风雪和沙尘等等。这种测试上的分类方法有什么问题，忽略了什么关键因素?

模拟测试无法真正模拟人类智力，特别是人类智力的不稳定演变。在实用场景下，跟无人汽车相关的人类智力演变开始阶段相对较慢；不稳定演变会发生在大规模应用后较长时间、普通人对无人汽车特点和缺陷了解比较透彻之后；不稳定演变会逐渐加快。太迟发现问题会造成严重后果。所以需要研究智力科学的基础理论和实验来寻找问题的本质。

AI: A Modern Approach 等等主流人工智能教科书无法分析人类特有智力的结构、系统及其不稳定演变，当然就缺乏对相关人工智能技术作正确测试的方法。而对人工智能技术和应用的不适当倡导也就有严重误导。大学教育误人子弟会有严重后果。

事实上，人类的教育模式、经济模式及思维和智力模式的不稳定发展和演变远比生态模式和基因的互动更加复杂。现有学术方法无法处理这种不稳定性，需要新的理论和方法来研究逻辑临界点的演变和发展以便处理这种不稳定演变。目前缺乏对 AlphaGo Zero 和各种后续系列以及自动驾驶汽车的逻辑临界点分析和实验。

Leukotomy 到底破坏了哪些人类特有智力，有没有破坏自由意志? Leukotomy 破坏的智力在无人汽车技术中是否都存在? 自由意志也是研究无人汽车技术能否适应未来模式不稳定演变的重要问题。这些问题需要更高层次的抽象思维。Leukotomy 如此严重错误七十多年不纠正，学术界和社会均已丧失诚信，也说明 selfish gene 能实现公益的理论错误。

哥德尔批评过逻辑实证主义（logical positivism 或 logical empiricism）的问题，这跟学术语言的使用规范有关。不正确的语言使用方法会严重误导科学实验的设计和对实验结果的解释，误导科学发展。这些在智力相关研究中正变得越来越严重，并危及科学和文明基础。

计算机是数学的一个子集。增强某些方面的数学功能本来就是研发计算机的主要目的。但数学不能覆盖一些重要的语言和科学功能，计算机也不能覆盖一些重要的数学功能。

Universal Approximation Theorem 和 Nash Embedding Theorems 适用条件有限制，各种计算机语言模型包括 word-embedding Vector Space 无法处理抽象思维、语义演变、高阶逻辑、诡辩分析和判断力等等问题。智力科学需要建立在严格控制条件的实验基础上。

Deepmind 缺乏让 AlphaGo Zero 及各后续系列参与严谨科学实验的公开性，所以笔者只能有条件地逐步公开实验细节，以保证这些人工智能实验的有效性，免受人类智力污染。

目前测试语言处理的各种数据集，缺乏科学和文明发展所需的关键性抽象思维、语义演变、高阶逻辑和诡辩分析等等能力的测试，明显高估人工智能而有很严重的误导。

基础科学研究不是市场经济，也不是计划经济。基础科学研究根本就不是经济，而是公共服务。把所有公共服务生硬地归入经济范畴，会导致许多弊端。滥用语言更有严重后果。

凯恩斯理论注重赤字经济会导致泡沫形成。小罗斯福政策则包含加强政府监管、防止市场炒作和投机。小罗斯福政策和凯恩斯主义在二战后至今日的各自成败需要更深研究。制造泡沫可带来 reward，但泡沫会出现不稳定演变而破灭，将社会带入陷阱。这也是 Deepmind 的 Reward is Enough 观点的问题。发现他们技术的问题，需根据智力科学理论有针对性设计实验。现在硅谷没有光刻机和芯片制造的优势，人工智能走错方向。风险投机家希望在泡沫破灭前准点退出、将损失留给别人，是否可行?

凯恩斯在 A Treatise on Probability 文中提出的观点在解释 causation 关系上也有问题。泡沫会破灭，概率也会发生模式不稳定演变。自由意志和各种人格及相应智力特征的关系，以及工具型人格的缺陷，也跟经济、教育和人工智能等等方面的研究密切相关。

笔者从可以设计实验的围棋和语言处理着手。语言、逻辑、causation 问题无法正确解决，司法、太空探索、医疗、教育、人工智能等等领域都会成为乱源。

科学和文明的延续和发展需要对从中学到大学的教育和研究作变革，建立和发展数和语言的智力结构分析，特别要对逻辑、公理系统和哥德尔不完备定理作新的理论解释和发展，为学生的思维和实践建立严谨的逻辑和科学基础。相关次序应该是:

1）完成智力科学理论和实验研究的主体部分;
2）变革前沿研究性大学的通识教育;
3）变革中学教育的体系和内容。

阿拉伯之春基本失败。美国最高法院文献中“Confucius taught that marriage lies at the foundation of government.”的错误背后，有重要逻辑和 causation 问题。美国大学的人文课程注重大量和广泛的阅读，确是优点。但读书不求甚解，也是明显问题。不喜欢批评是人之常情，所以压制正确意见很容易。但听不到正确意见，问题就大了。

斯坦因先生（Marc Aurel Stein）算是把中国文化带到西方。司徒雷登先生（John Leighton Stuart）则把西方文化带到中国。他们在中国都曾经如鱼得水。但他们对中国文化的了解仅在实用层次。停留在这种层次的理解不够。等大问题出来，就来不及了。人类智力模式的高度不稳定演变，造成司徒雷登先生在中国境遇上的剧变。这里有教育上的问题。

前沿研究性大学应把更严谨有效的 causation 理论、逻辑判断方法和思维模式演变作为研究和教育的核心基础。应警惕对科学严谨性的厌烦和智力的退化，免受短期利益误导。

哥白尼受嘲弄，开普勒死于贫困，伽利略受迫害。前几年有人想抬高第谷，哥白尼、开普勒、伽利略是不是又开始令人厌烦了？当年第谷得到很多赞助，但第谷的学术体系是错误的。他的天文观察仍有价值是侥幸，因为当时争论各方对天文观察方法没有争议。

而智力研究中什么是有效观察量有重要争议。Demis Hassabis 和很多人工智能专家有方向性错误。人工智能的局部进展价值有限，远远弥补不了学术体系的方向性误导。

现在全球性生产过剩已经出现，世界正在剧变。贸易战和新冠病毒疫情造成的物资短缺是暂时假象。货币宽松和泡沫无法持久。掩盖严重学术错误，让公众受愚弄，高房价、高收入还能维持多久？笔者的学术批评尖锐，是因为有人歪曲笔者观点、淡化和模糊问题。

探索新科学需要平等讨论、公正实验，跟从魔鬼那里获取秘诀的浮士德有本质不同。需要不断接受更严谨实验的挑战。不要迷信智商之类没有科学基础的观念。智力不是 Totally Ordered Set，其结构和特征也需要进一步的实验研究。这里的关键是结构，不是计算。

以下附录中，笔者分析了一些对科学和文明较为重要的问题。有些是重要逻辑临界点，可作证伪的反例。以此展开研究，可设计严格控制条件的实验验证各种智力科学原理。

附录一、历史和思维模式研究的一个案例

人类是一种特殊的动物。人性中包含了动物性。智力研究也跟人性及其复杂性有关。

由于心理学上的 Hawthorne effect 等等问题，思维模式研究极为困难。而历史资料可以避免研究者和被研究者之间的干扰。这里介绍一个通过历史资料研究思维模式的实例。

韩日之间有慰安妇问题。笔者缺乏必要资料，无法对此历史问题作出判断。

不过笔者在了解雾社事件的历史背景资料时，发现日治时代的台湾有联姻政策，鼓励日本警察跟原住民头目之女联姻。甚至有资料认为此联姻政策很多以悲剧收场，因为被日本政府强迫结婚的日籍警察抛弃原住民妇女。

如果这方面资料属实，则帝国时代的日本政府在两性关系上有强迫性的思维模式；联姻政策名为鼓励，实际上对日籍警察有强迫性的，有命令日籍警察执行的情况。

而研究当时的联姻政策在实施中是否有强迫性，则比较容易。本案例仅根据一些线索的分析从新的角度提出研究建议。笔者并未对问题作出结论。

人在乱世，非常悲哀，女性尤其如此。这个问题是否值得研究?

附录二、语义演变和复杂性分析的案例

先介绍一些对人和社会及其演变的重要语言描述。这些对准确理解各种司法体系的异同至关重要。以此为 context 可以设计一些问题来测试和区分人类和人工智能的语言智力。

《管子》分析了皇道、帝道、王道、霸道四种模式。嬴政试图融合皇道和帝道，创立皇帝名位。自此皇和帝的语义已有新发展，不同于先秦的皇道和帝道。这些模式下的司法体系也是基于不同的原则。

项羽、韩信试图恢复王道。项羽先前的诸侯上将军接近霸道，后来的霸王接近王道中的天子。韩信想做的王其实是王道中的诸侯。但两人都失败。刘邦则在皇帝模式和王模式之间妥协，保留同姓诸侯。

汉宣帝说“霸王道杂之”和《管子》中“明一者皇，察道者帝，通德者王，谋得兵胜者霸。”的王、霸语义既有关联又明显不同。前者指皇帝模式下两种思维、运作方式，包括皇帝驾驭臣下的方法。后者指不同于皇帝模式的两种模式。前者从后者演变而来。

汉元帝正逢盛世，未必真能理解世道的险恶。汉宣帝有微时故剑的典故，其中主角许皇后，即汉元帝之母，被权势毒杀。汉宣帝前任汉昭帝虽是明君，却生活在权势的阴影和恐惧中，很短命。汉元帝笃信儒家的仁、义、礼、智、信，但汉元帝之后有王莽篡权。再后来东汉也有很多皇帝是傀儡且短命。所以儒家的概念体系也无法概括人和社会的复杂性。

事实上，真实世界的复杂性，远远超出语言的复杂性，所以各种语言都无法完整描述真实的人和社会演变及其问题，只能近似地描述部分特征。这跟纯数学不同。理解语言背后的人类智力则更加复杂和困难。所以，需要不断探索和发展新的语言和理论。

皇、帝、王不同于英文 emperor、king 或 queen。称 Queen Victoria 为英国女皇并不准确，实际上她是 Queen of the United Kingdom(UK) 和 Empress of India，这是两个不同职位，对应不同法律职能。UK 逐渐走向君主立宪；英属印度帝国实际上在 UK 的管理下，并非民主国家。而 UK 宪法并非一部法律文件，其包含惯例和共识，有非语言成分，需要语言之上的人类智力理解。这些对理解 UK、英属香港、英属印度复杂且不同的司法体系非常重要。

中文沙皇也不准确。沙皇从 Царь 来。Царь 起源于拉丁文 Caesar，即凯撒，古罗马统治者，其概念发展到 Царь 已有很大的演变。俄罗斯帝国君主 Царь 和 Император 两个头衔，也有不同法律职能。不但不同于中国君主概念，也不同于英国君主概念，其法律体系也不同。

美国宪法语义，由美国最高法院法官解释。法律由人类制定，为人类服务，由人类解释、裁决。如果不采用非法手段，人工智能不可能在宪法解释能力上取代最高法院法官。人工智能也不可能如 Demis Hassabis 所说 potentially a meta-solution to any problem。研究人类智力的结构特征，研究类人智力，可以帮助发现和理解人工智能语言处理的很多问题。

现代商业竞争跟霸道有相似之处。霸道持续时间较短，公司的商业霸权也很少超过百年。

从思维模式看，瑞士和新加坡都注重理性，却是有些不同的理性，对应的司法体系也不同。而太空探索和发展智力新科学需要更加不同的理性和更加不同的司法体系。

NIH 目前的研究方法，无法理解生命科学和智力科学的不同，也无法正确测试 leukotomy、基因编辑和太空环境中人类快速演化对智力的影响，更无法研究哪些变异会有利于人类在太空的发展，哪些会造成损害。所以需要更加准确、全面、科学的智力测试理论和方法。

由于缺乏准确的智力系统结构特征分析，与 leukotomy 相关的统计数据没有选择正确、全面的观察量。所以这些统计结果根本不能反映手术对智力的真实影响。

根据各种语义分析案例，笔者可以设计很多语言测试的方法，来测试那些对文明和科学发展非常重要却被现在的语言测试集忽视的智力问题，包括测试人工智能的智力。本案例涉及历史知识。而更广泛的测试既需要各种专题的问题，也需要较少涉及专门知识的问题。

类似抽象了静止和运动的惯性及哈密顿量等等概念形成了物理科学的基础，智力科学也需要许多新概念来有效设计实验、发现规律。根据这些概念，也可针对很多领域设计测试。

附录三、关于科学验证方法的退化

由于很多物理科学前沿研究问题越来越难以验证，近年科学界有人主张证明的概念也需要演化。但他们提出的新验证方法实际上是退化版的，而不是进化版的，典型的表达如：

> First, if scientists have tried, and failed, to come up with an alternative theory that explains a phenomenon well, that counts as evidence in favor of the original theory. Second, if a theory keeps seeming like a better idea the more you study it, that’s another plus-one. And if a line of thought produced a theory that evidence later supported, chances are it will again.

这种方法并没有依严谨的逻辑去证伪其它可能来真正证明一理论，明显达不到科学验证的标准。自然研究经历了自然历史（Historia Naturalis）、自然哲学（Philosophiae Naturalis）、自然科学（scientia naturalis）的发展。这种退化版的方法仅属于自然历史范畴的研究。

这也是物理科学研究正遭遇越来越严重的瓶颈，科学思维的严谨性正在退化的证据。所以需要智力新科学来研究这种现象，并通过严谨的智力科学实验带动科学严谨性的继续发展。

附录四、对多重宇宙和量子力学铨释争议的智力结构分析和语言处理

智力科学的另外一个重要应用是分析多重宇宙和量子力学铨释理论的性质。基于相同的科学实验结果，出现许多不同量子力学铨释，就涉及智力科学问题，涉及科学逻辑的演绎基础及有效范围。如多世界诠释是否过度解释，不同宇宙按定义是否可相互测试？如可相互测试，不就仍在同一个宇宙中？

学术界广泛使用过度解释方法，如数学即是过度解释的产物。过度解释的理论可能无法证伪，虽不等于错误，仍可能带来问题，所以需要智力新科学识别过度解释并对数、语言和各种理论、方法作智力结构和复杂性分析，规范过度解释的方法包括数学等等的使用。

多世界诠释理论存在争议。有该理论的辩护者声称多世界主张不是该理论的假设，而是该理论的推论。按这种表达，这个多世界推论仍然有无法测试的问题。

另有最近文章Are there really many worlds in the "Many-worlds interpretation" of Quantum Mechanics?认为: the development of «decoherence theory» revealed that, using the standard formalism of quantum mechanics, macroscopically distinct branches of the wavefunction were almost entirely free from interference and evolve approximately classically。该观点使用了almost一词，是否已放弃多世界主张，不管是假设还是推论？

有些问题是由于多世界等等语义有演变造成混淆。减少混淆的最好方法是多世界诠释（The Many-worlds Interpretation）名称应该仅仅保留给严格意义上的多世界主张，即按严格定义无法相互观察或测量的多世界。这样的多世界主张实际上也是过度解释的。而对其它各种量子力学诠释也应该进一步分析其可测试性以及是否过度解释等等问题。

科学有测量极限，引入各种新理论也难以真正解决量子力学的诠释争议。前述那种退化版的“证明”方法不是真正的证明，当然更加无法解决这些诠释争议。这些问题使量子力学语义高度不稳定，造成信息混沌和熵增加现象。需要智力新科学带来更高境界的理性。

笔者在本附录开头对多世界诠释问题的分析，没有参考和引用物理学理论的细节，仅用一句话“如可相互测试，不就仍在同一个宇宙中？”完成逻辑推导。不但可得出与众多物理学家相同的结论，而且笔者的推理不必依赖于有争议的物理理论，且可对各种诠释的性质作进一步澄清。这种更高抽象程度的简洁逻辑可以推广到其它学术研究。

使用提升抽象层次等等语义方法和其它智力科学方法，虽无法取代物理和生命科学研究，但可解决一些特殊问题，包括一些物理和生命科学难以解决、甚至无法解决的问题。

计算机模型是数学的子集，也有过度解释的问题。用数学或计算机模型表达自然语言和科学语言，会有本质问题，智力新科学需要新的方法识别过度解释。

附录五、计算思维、人工智能和科学研究的区别

研究智力科学，最好回顾一下物理科学和生命科学的研究。

除了达尔文，华莱士也对演化论作了重要贡献。不过华莱士的理论跟达尔文的理论有些不同之处。据说华莱士受了马尔萨斯的影响。而达尔文受计算思维的影响较少。

南美蝴蝶振动翅膀，可能在非洲引起风暴。但对蝴蝶来说，这纯属巧合。蝴蝶并不知道能否及在何处引起多大风暴。诺伊曼控制气象失败，现在仍只能改变某些小气象。当然，空调房内控制气象和电梯自动驾驶早已实现。这些显示了计算思维的应用价值及局限性。

1975年美国在越南战争中失败，标志着Robert McNamara式精英和概率统计方法的失败。而1976年Steve Jobs和其他人建立Apple、开发面向大众的计算机，标志着逻辑的胜利和新的精英时代。以一阶逻辑为基础的计算机技术在很多方面有应用价值。

但一阶逻辑有局限性。人类智力远比气象复杂。人工智能和基于一阶逻辑的传统计算机应用也有本质不同。几十年后计算机试图进入无人汽车和完全的自然语言处理，已是过度扩张。从炒作人工智能和量子计算的迫不及待看，现在计算机精英们恐怕找不到成熟的科技积累来开创新时代。量子计算无法解决一阶逻辑的局限性。

Deep Learning和reinforcement learning实际上是迷信产生的根源。缸中之脑（Brain in a vat）无法区分数据集中的全称判断和真实世界中的全称判断，导致迷信。所以科学需要更严谨的逻辑和实验规范。

Demis Hassabis宣称他们的技术potentially a meta-solution to any problem，就是全称判断导致的迷信。有人工智能专家说“人类正在扮演上帝的角色”也是迷信，哥德尔定理证明了metaphysics from human does not work。

这里的迷信指哲学和科学上的迷信，语义不同于英文superstition。超弦理论、量子力学诠释争议、多重宇宙等等物理科学的前沿问题，使超自然的神在科学上反而难以证伪。所以科学无法介入很多宗教问题，笔者也不跨越科学和宗教的边界。

无人汽车能收集到的数据，只是真实世界的极小部分，丢失了真实世界未来的模式不稳定演变。笔者可以在不知道无人汽车程序设计的情况下找出其一小部分问

题。这意味着未来非精英阶层民众在无人汽车大规模应用较长时间后，也可以在不知道无人汽车程序设计的情况下，找出更多精英们无法找出的问题。

对很多人来说，计算机算力已经过剩。5G/6G、Starlink 或电动汽车只是小改进。很多更加重要的问题，却无法用计算机解决。几百年的大周期是不是快要来临?

科学研究不是按赞成者多少来决定对错，而只能靠实验来验证。科学不能证实，只能证伪。如果无法证伪，就只能靠不断地公开接受更严谨实验的挑战来进步。科学多数派应该抑制自己人数上的优势，让少数派有机会发表意见、设计实验，避免认知上的错觉。

附录六、关于科学、经济和管理的研究

近日网上看到沙烨谈论人工智能、经济发展和财富分配等等。

沙烨这些观点肯定不是笔者教的。笔者有理由认为图灵测试无效且有误导，智能实验需要严谨的逻辑、因果分析和科学规范。笔者并不研究所有经济问题。而跟科学探索相关的经济研究，似乎沙烨做得不够，却是笔者的关注重点。

马克思经济理论的重要基础是社会化大生产可以带来规模优势，利用这种规模优势可以给工人支付更高薪水、带来更大购买力和需求。马克思这部分经济理论跟汽车企业家老福特的经营实践有相似之处。

老福特通过生产流程的改进，大大提高了生产效率和规模。他的技术流程和高薪政策也带动了社会生产率的提高和高薪化以及更大的购买力。他的模式为社会化大生产的规模优势提供了一种较好的实现方式。老福特的成功是在当时的工业化人口大国美国取得的。这种成功经验也和最近几十年中国利用人口规模优势发展成制造业大国的历史吻合。

不过马克思没有研究社会化大生产的思维模式是否对科学研究，特别是对探索新科学，有某些不利影响。当然各种自由经济理论也没有对此作很好的研究。而社会化大生产的庞大管理体系和市场都很容易受权势的操纵和误导。另外，以色列的基布兹似乎有经济以外的价值。智力新科学应该研究这些问题，为教育和研究的多样化提供理论基础和结构启示。

事实上，老福特的个人专断，就把他的汽车龙头企业带入困境。福特二世引入健全的管理体系，则拯救了福特公司。在美国汽车三巨头时代，通用汽车公司以规模占优；第二位的福特公司有点危机意识，也有较好的文化和管理；克莱斯勒公司规模较小，苦苦挣扎。机灵的艾科卡虽拯救了克莱斯勒，但也只能靠不断合并来扩大规模、继续生存。

老福特式的高薪经营模式在产业的新兴阶段具技术优势时常常会有效。但在产品变得较成熟，尤其是变成 commodity 以后，利润率大大下降，企业家无法支付高薪水，就不得不寻找更低成本的劳动力。底特律的衰落就是一例。

日本、韩国、中国汽车业的兴起背后，既有较低成本的因素，也有政府的助力。传统汽车业和政府都有庞大的管理体系，在文化上既比较接近，又有互补性。威权体系延续下来的文化传统在学习比较成熟的技术和产品制造上较有效率和优势。

而电动汽车大大简化了汽车结构，也大大减小了汽车业雇员规模、简化了管理复杂性。无人汽车则是完全不同的问题。汽车的机械、材料、电子等等设计都有成熟的物理科学作基础。而对人类智力的理解，既没有科学基础，现在产业界也不愿接受科学新实验的挑战。这使科学的研究方法和实验规范在伽利略之后又重新成为问题。

不断探索新科学才能不断造就新产业、推动经济的继续发展。但探索新科学可能在很长时间内无法带来经济价值，甚至没有社会影响。所以，其所需要的资源往往无法靠市场规律获取。而政府研究基金在探索新科学上有没有走对方向，就变得至关重要。

Robert McNamara 等等 Whiz Kids，协助福特二世复兴了福特公司。McNamara 以后出任美国国防部长，主导越南战争；其后又任世界银行行长。Henry Kissinger 在 McNamara 之后，也对越南战争有关键影响。阿富汗前总统阿什拉夫·加尼则是哈佛大学、欧洲工商管理学院和斯坦福大学商学院为世界银行设立的领导培训计划选拔的精英。

据网上信息，McNamara 和 Kissinger 也是政策研究机构 Aspen Institute 的特别顾问。两人处理不同文化的人类智力时都出现严重失误。笔者不知道授予 Henry Kissinger 诺贝尔和平奖的决定是否正确，因为一起获奖的黎德寿没有领奖，而据说五名评委中有两名因此辞职。加尼既有阿富汗背景，又有广泛国际经历，也失败。所以，人类智力研究在人工智能技术和经济、管理、教育等等研究上都是严峻挑战。

研究者不应该迷信小范围的成功经验，而需要引入严谨的科学新方法。科学方法既不同于权威体系和官僚制度，也不同于市场规律和民主政体，不能以钱和选票来决定科学对错。所以，科学方法不同于社会主流制度，这在多样化管理方面形成特别的困难。

全球化浪潮及其失败显示了现在的地球生态圈作为整体其实仍相当封闭，熵定律的影响在增强。而物理科学研究已相对成熟，太空探索和发展的主要困难在于生命科学和智力科学。联合国成立于 1945 年。《外层空间条约》于 1967 年生效，其语言和逻辑问题涉及更广阔、深远的未来，有待更严谨的研究。笔者并不教人曲学阿世。

机器人 Pepper 的失败说明人工智能连一些基础智力原理仍没有理解。李飞飞教授的“以人为本”人工智能研究能不能解释清楚问题的关键在哪里？人类是否有自由意志，Pepper 这样的角色是否又需要自由意志，自由意志跟随机性又有什么区别？

而 AlphaGo Zero 是否有自由意志，也是能否达到 superhuman 的关键之一。无人汽车技术是否有自由意志、科学研究是否需要自由意志，则是更加重要的问题。

最近有一篇文章 Deep Learning Is Hitting a Wall（https://nautil.us/deep-learning-is-hitting-a-wall-14467/）。如果该文及其引用的学术论文内容属实，主流学术界似乎已经开始认识到他们的错误。但他们只能发现一小部分问题，甚至无法发现未来模式不稳定演变下新出现的大概率问题。那种问题需要有针对性地有效控制条件的实验才能发现。

用 not scaling 和 toxicity 等语言来描述问题，说明他们不懂人类智力的结构。由于他们缺乏适当的语言和概念来研究和找出人类特有智力的结构特征，也就无法比较人工智能和人类特有智力的本质区别，当然也无法找出问题的关键，更无法有针对性地设计出实验的控制条件。所以，他们虽然强调测试的重要，却并不真正知道如何设计好的测试。

而现在大学里跟智力有关的很多教育也是错误的。智力研究上的错误，并非仅仅影响人工智能技术发展，也误导教育、经济、司法、医疗等等方面的发展，会造成文明的衰退。缺乏对人类智力及其发展的正确理解，未来太空时代人类的快速演化，会导致更严重后果。

附录七、美亚高科技风险投资的一些案例和问题

因缺乏研究资源，笔者只能抛砖引玉。下列几个案例有些重要特征值得深入研究：

1. 陆宏亮：在无线通讯的大趋势下，针对成熟方案尚难以覆盖的区域，UTStarcom 研发暂时性的替代方案，并非最新尖端科技。

2. 张首晟：跟踪最新科技，但投资的往往不是该领域最强团队。作为投资人的控制能力也有争议。

3. 朱敏：本业 WebEx、Zoom 成功。跟其他投资人形成的广泛合伙人关系似有泡沫。对其它行业投资有可疑资本运作手段和对被投资团队不适当操控的争议。有成为失信人报道。

4. 孙正义：以电讯业起家。管理能力较强。有高科技战略性投资，包括收购 Aldebaran Robotics 和 ARM、开发机器人 Pepper，但效果存疑。

华为智能手机本来有力压群雄之势，后受毁灭性打击。Nvidia 收购 ARM 也受挫。Alphabet 的无人汽车和量子计算项目出现泡沫，无人汽车有巨大隐患。太空探索的主要挑战在于生命科学和智力科学。所以，纠正学术错误、探索新科学，才是正道。

再提醒一下：哥白尼受嘲弄，开普勒死于贫困，伽利略受迫害。伽利略抽象了静止和运动的本质而建立起来的新科学，徐光启没有学，康熙没有学，伍秉鉴有没有学？

瓷器、丝绸、茶叶、香料、银矿的实用价值很大。但三十年战争和明清战争使这些勉强的繁荣化为泡影。欧洲在三十年战争中期开始科学革命。清朝的康乾时代没有强劲竞争对手，也基本不依赖外来技术；但随后就进入了中国千年大衰退的末期和谷底。

现在科学革命已近四百年，如果严重的学术错误不纠正，新科学探索无法取得重要进展，文明很快就会衰退。

在外太空，人类无法直接生存，生物物种包括人类甚至会出现剧烈演化。无论是维京人的冒险模式，还是哥伦布、麦哲伦的冒险模式，在太空探索和发展上都会无效。

附录八、关于自然历史、自然哲学、自然科学

哥伦布不懂新科学，也能发现新大陆。他所依靠的是新的地理理论。哥伦布在计算上是错误的，因为美洲不是印度；但在结构上部分正确：事实上地球确实接近球形。这部分的正确，已经可以为冒险家的成功奠定基础了。哥伦布无法知道他自己的对错。麦哲伦船队环球航行后才能验证地球是球形的、美洲不是印度，这在哥伦布去世后十几年才发生。

地理理论经过验证和修正变成常识（common sense），但不是科学理论，因为地球结构会演变，虽然演变非常缓慢。大陆会漂移，各大洲相对位置会改变；而地球本身也可能瓦解，失去球的形状。这些是地质学和天体学的研究范围，是自然历史（Historia Naturalis）。

而人类智力结构的演变非常快速。对人类智力研究，常识和自然历史方法会失效。所以，笔者设计可控条件的实验来验证一些智力原理，以此可以提前分析人工智能技术的结构问题。现在计算机围棋的对弈、无人汽车的测试以及其它智能测试都是自然历史（Philosophiae Naturalis）方法，不是科学实验和科学研究方法，缺乏可靠性。

对智力科学来说，区分常识和科学理论，区分自然历史（Philosophiae Naturalis）、自然哲学（Philosophiae Naturalis）、自然科学（scientia naturalis）的结构，非常重要。

附录九、关于人工智能教科书、类人智力和类人机器人

神经科学仅在生理层次的动物性智力：视觉、声觉、运动、情绪、记忆、经验、控制和自我意识等等方面有进展，无法真正研究人类特有智力如语言、数学、哲学、科学的发展。

Marvin Minsky 教授的著作 The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind，把情绪（emotion）当成人类智力的关键。可惜他的观点错误。情绪只是动物性智力，人类也非机器（machine）。这个问题在各种人工智能教科书中也存在。事实上，计算机可能使用一些情绪性技巧，比如装傻、装糊涂、胡搅蛮缠等等，而在 Turing Test 中蒙混过关。分析、验证这个问题，首先需要对人类特有智力和机器的本质特征作准确的结构性抽象和比较。这也再次显示了语言是人类特有智力的重要基础。

尽管 Minsky 教授征求别人对该书的看法，学术界并没有对他的观点作深入探讨和证伪。这也影响到后来教科书的准确性。笔者曾想把自己看法告诉他，却不幸发现他已去世。

主流教科书 AI: A Modern Approach 第三版开始，出现很多严重错误。特别是对人类智力的重要结构、系统特征和数学、计算机技术的本质缺陷缺乏理解，因此缺乏准确、有效测试人工智能技术的方法，却对不成熟的人工智能技术和应用盲目倡导。这些问题直接造成 SAE level 4 自动驾驶测试、验证的条件不稳定、不准确。问题无法纠正，则说明现在人工智能研究领域缺乏必要的平等讨论和公正实验的科学规范。

机器人 Pepper 被定位成能理解情绪智力的机器人。这种定位当然无法抓住人类智力的关键结构特征。类人机器人的研究，最好从自由意志在人类智力结构中的作用着手。而自由意志的研究，又可以从笛卡尔的哲学、贝多芬的音乐和哥德尔的数学理论着手。

类人机器人研究，远比想象复杂，从中可以发现机器的缺陷。其各项智能可以当作人工智能技术的试金石。需要进一步从哲学、数学和科学上分析研究人类特有

智力的结构特征；类人机器人的发展上限；和类人机器人在未来，特别在太空探索发展中可发挥什么价值。

意识（consciousness）也是认知研究中经常引起混淆和误导的概念，会把很多不同问题混在一起、无法理清，不是有效观察量。对某些声称跟意识有关的实验，应该根据不同情况分别使用自我意识或其它更准确的概念。自我意识也是动物性智力，并非人类特有智力。

顺便提一下，杀戮（killing）也是动物性智力。所以研制人工智能武器并不一定需要人类特有智力，笔者也不研究人工智能武器。

科学研究也远比辩论复杂，还需要设计有效控制条件的实验以及对实验结果作有效的分析和解释等等。这些需要更多的人类特有智力特征。

生命科学只能研究演化，而智力有进化和退化的区别。控制相对容易，发展文明和科学比较困难；因为控制是动物性智力，而发展文明和科学需要人类特有智力。事实上对人类的不适当控制会导致人类智力的退化和丛林法则的泛滥。

应该研究自由意志在希帕索斯、阿那克萨哥拉、伽利略等人推动学术和文明进步的研究中所起的作用。对希帕索斯的打击和对伽利略的控制、训练和惩罚并无法阻挡新的学术和科学的传播，说明当时文明处在上升阶段。而在这两个历史事件之间，古希腊的日心说则被遗忘一千多年。

附录十、关于因果作用和纳什均衡（**Nash equilibrium**）

因果关系不同于逻辑关系，这是类人智力研究和未来太空发展的关键之一。

在逻辑上，广义相对论已不需要日心说；但在因果关系上，日心说对人类发展出广义相对论却必不可少。没有日心说，就无法形成力学基础和经典物理学，当然就不会出现广义相对论。而驱动人类从远古天文观念到日心说，再到广义相对论的因果作用是什么，能否在类人机器人中实现，则是智力科学研究的重要问题。

在数学上是否有更优解存在，跟 deep learning、reinforcement learning 或其它人工智能方法和计算方法能否找到这个更优解，是两回事情。数学上有种构造主义（Constructivism），不过构造性数学的研究至今并不成功。所以智力科学研究需要新的方法。

纳什均衡是一种纯数学模型，在物理科学中也许能找到某些应用价值，但在智力科学中并非准确的模型。当人类成为纳什均衡中的囚徒，人的价值受到践踏和浪费，制定纳什均衡规则的势力会出现衰退，迟早会无法维持规则的有序性。或者他们会受其他人挑战，或者人类作为整体会衰败。人类未必是适应得最好的生物，在外太空中尤其如此。这里的因果作用关系，包括自然选择，远比人类能够用数学表达的更加复杂。

主流教科书 AI: A Modern Approach 不理解哥德尔不完备定理，不了解人工智能技术的逻辑基础问题，更不了解对人工智能技术有效性至关重要的因果作用，所以缺乏有效测试理论和方法。其 Wind tunnel approach 误导在于混淆了物理科学和智力科学的本质区别。

而研究 Alan Turing 的思维模式中缺少了什么，John Nash 的思维模式中缺少了什么，Demis Hassabis 的思维模式中又缺少了什么，对教育和智力科学研究有相当价值。当然这些研究是纯学术研究，是比较不同思维模式的差异，跟未来智力科学发展有关，并非指责上述学者，因为人无完人。当然知错却不纠正应受指责。

教科书问题会误人子弟，比较紧急。而没有二、三十年资源充足的研究，探索新科学很难产生直接应用价值。所以，一开始只有修正教科书和测试人工智能技术等等的间接价值。

现在，笔者的智力新科学研究受压制、排斥、打击已有十年，已整整浪费了十年。